MULTISENSORY SPEECH DETECTION

CROSS-REFERENCE TO RELATED APPLICATION

This patent application is a continuation of, and claims priority under 35 U.S.C. § 120 from, U.S. patent application Ser. No. 16/017,580, filed on Jun. 25, 2018, which is a continuation of U.S. patent application Ser. No. 14/753,904, filed on Jun. 29, 2015, which is a continuation of U.S. patent application Ser. No. 14/645,802, filed on Mar. 12, 2015, which is a continuation of U.S. patent application Ser. No. 12/615,583, filed on Nov. 10, 2009, which claims priority under 35 U.S.C. § 119(e) to U.S. Provisional Application 61/113,061, filed on Nov. 10, 2008. The disclosures of these prior applications are considered part of the disclosure of this application and are hereby incorporated by reference in their entireties.

TECHNICAL FIELD

This instant specification relates to speech detection.

BACKGROUND

As computer processors have decreased in size and expense, mobile computing devices have become increasingly widespread. Designed to be portable, many mobile computing devices are lightweight and small enough to be worn or carried in a pocket or handbag. However, the portability of modern mobile computing devices comes at a price: today's mobile computing devices often incorporate small input devices to reduce the size and weight of the device. For example, many current mobile devices include small keyboards that many people (especially those with poor dexterity) find difficult to use.

Some mobile computing devices address this problem by allowing a user to interact with the device using speech. For example, a user can place a call to someone in his contact list by simply speaking a voice command (e.g., "call") and the name of the person into the phone. However, speech can be difficult to distinguish from background noise in some environments, and it can hard to capture user speech in a manner that is natural to the user. In addition, it can be challenging to begin recording speech at the right time. For example, if recording begins after the user has started speaking the resulting recording may not include all of the user's voice command. Furthermore, a user may be notified that a spoken command was not recognized by the device after the user has spoken, which can be frustrating for users.

SUMMARY

In general, this document describes systems and techniques for detecting speech. In some implementations, a mobile computing device can determine whether a user is speaking (or is about to speak) to the device based on the changing orientation (i.e., distance from or proximity to a user and/or angle) of the device. For example, the device may use one or more sensors to determine if the user has made a particular gesture with the device such as bringing it from in front of the user's face to a normal talk position with the device at the user's ear. If the gesture is detected, the device may emit a sound to indicate that the user may start speaking and audio recording may commence. A second gesture of moving the device away from the user's ear can be used as a trigger to cease recording.

In addition, the device may determine whether it is in a specified "pose" that corresponds to a mode of interacting with the device. When the device is placed into a predefined pose, the device may begin sound recording. Once the device has been removed from the pose, sound recording may cease. In some cases, auditory, tactile, or visual feedback (or a combination of the three) may be given to indicate that the device has either started or stopped recording.

In one implementation, a computer-implemented method of multisensory speech detection is disclosed. The method comprises determining an orientation of a mobile device and determining an operating mode of the mobile device based on the orientation of the mobile device. The method further includes identifying speech detection parameters that specify when speech detection begins or ends based on the detected operating mode and detecting speech from the user of the mobile device based on the speech detection parameters.

In some aspects, detecting an orientation of a mobile device further comprises detecting an angle of the mobile device. In yet further aspects, detecting an orientation of a mobile device further comprises detecting a proximity of the mobile device to the user of the mobile device. Also, determining an operating mode of a mobile device comprises using a Bayesian network to identify a movement of the mobile device.

In another implementation, a system for multisensory speech detection is disclosed. The system can include one or more computers having at least one sensor that detects an orientation of a mobile device relative to a user of the mobile device. The system can further include a pose identifier that identifies a pose of the mobile device based on the detected orientation of the mobile device. In addition, the system may include a speech endpointer that identifies selected speech detection parameters that specify when speech detection begins or ends.

In certain aspects, the system can include an accelerometer. The system can also include a proximity sensor. In addition, the system may also include a gesture classifier that classifies movements of the mobile device.

The systems and techniques described here may provide one or more of the following advantages. First, a system can allow a user to interact with a mobile device in a natural manner. Second, recorded audio may have a higher signal-to-noise ratio. Third, a system can record speech without clipping the speech. Fourth, a system may provide feedback regarding audio signal quality before a user begins speaking. The details of one or more embodiments of the multisensory speech detection feature are set forth in the accompanying drawings and the description below. Other features and advantages of the multisensory speech detection feature will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. **1** is a conceptual diagram of an example of multisensory speech detection.

FIG. **2** is a block diagram of an example multisensory speech detection system.

FIG. **3** illustrates an example process of multisensory speech detection.

FIG. **4** illustrates an example alternative process of multisensory speech detection.

FIGS. **5**A and **5**B illustrate coordinate systems for gesture recognition.